Снижение размерности пространства в задачах декодирования сигналов

Исаченко Роман Владимирович

Диссертация на соискание ученой степени
кандидата физико-математических наук

05.13.17 – Теоретические основы информатики

Научный руководитель: д.ф.-м.н. В. В. Стрижов

Москва, 2021 г.

Снижение размерности пространства в задачах декодирования сигналов

Цель

Исследовать зависимости в пространствах объектов и ответов и построить устойчивую модель декодирования временных рядов в случае коррелированного описания объекта исследования.

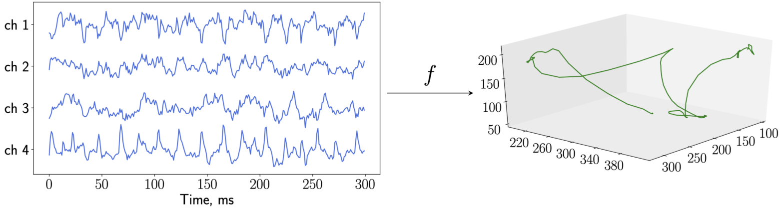
Проблема

Целевая переменная – вектор, компоненты которого являются зависимыми. Требуется построить модель, адекватно описывающую как пространство объектов так и пространство ответов при наблюдаемой мультикорреляции в обоих пространствах высокой размерности.

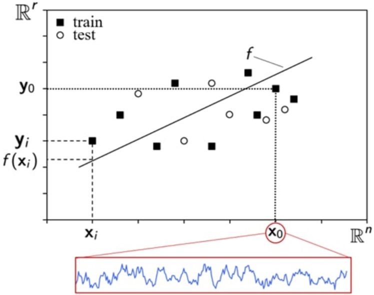
Решение

Для учёта зависимостей в пространствах объектов и ответов предлагается снизить размерность, используя скрытое пространство. Предлагаются линейные и нейросетевые методы согласования связанных моделей в пространствах высокой размерности.

Задача декодирования сигналов



Регрессионная модель декодирования



Проекция в скрытое пространство

$$
\begin{tikzcd}
\mathbf{X} \subset \mathbb{R}^n \arrow[rr, "\mathbf{f}"] \arrow[dr, "\mathbf{P}"'] & & \mathbf{Y} \subset \mathbb{R}^r \arrow[dl, "\mathbf{Q}"] \\
& \mathbf{T}, \mathbf{U} \subset \mathbb{R}^l &
\end{tikzcd}
$$

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathsf{T}} + \mathbf{E}_\mathbf{x}$$

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}^{\mathsf{T}} + \mathbf{E}_\mathbf{y}$$

$$\operatorname{cov}(\mathbf{T}, \mathbf{U}) \to \max_{\mathbf{P}, \mathbf{Q}}$$

Задача регрессионного декодирования

$\mathbf{y} = f(\mathbf{x}, \mathbf{\Theta}) + \varepsilon = \mathbf{\Theta x} + \varepsilon$ – модель, $\mathbf{\Theta} \in \mathbb{R}^{r \times n}$ – параметры модели.

Функция потерь линейной модели декодирования

$$\mathcal{L}(f, \mathbf{X}, \mathbf{Y}) = \Big\| \underset{m \times r}{\mathbf{Y}} - \underset{m \times n}{\mathbf{X}} \cdot \underset{r \times n}{\mathbf{\Theta}}^{\mathsf{T}} \Big\|_2^2 \to \min_{\mathbf{\Theta}}; \quad \mathbf{\Theta}^{\mathsf{T}} = (\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{Y}.$$

Для устранения сильной линейной зависимости столбцов матрицы $\mathbf{X}$ предлагается использовать методы снижения размерности пространства.

Метод проекции в скрытое пространство

$$\mathbf{X} = \mathbf{T} \cdot \mathbf{P}^{\mathsf{T}} + \mathbf{E}_{\mathbf{x}} = \sum_{k=1}^{\ell} \mathbf{t}_k \cdot \mathbf{p}_k^{\mathsf{T}} + \mathbf{E}_{\mathbf{x}},$$

$$\mathbf{Y} = \mathbf{U} \cdot \mathbf{Q}^{\mathsf{T}} + \mathbf{E}_{\mathbf{y}} = \sum_{k=1}^{\ell} \mathbf{u}_k \cdot \mathbf{q}_k^{\mathsf{T}} + \mathbf{E}_{\mathbf{y}}.$$

$$\hat{\mathbf{Y}} = f(\mathbf{X}, \mathbf{\Theta}) = \mathbf{X\Theta}; \quad \mathbf{U} \approx \mathbf{TB}, \quad \mathbf{B} = \mathrm{diag}(\beta_k), \quad \beta_k = \mathbf{u}_k^{\mathsf{T}}\mathbf{t}_k/(\mathbf{t}_k^{\mathsf{T}}\mathbf{t}_k).$$

Метод проекции в скрытое пространство

Утверждение (Исаченко, 2017)

Вычисленные вектора $\mathbf{t}_k$ и $\mathbf{u}_k$ с помощью итеративной процедуры обновления:

$$\mathbf{t}_k := \frac{\mathbf{X}_k \mathbf{w}_k}{\|\mathbf{w}_k\|}; \quad \mathbf{w}_k := \mathbf{X}_k^{\mathsf{T}} \mathbf{u}_{k-1} / (\mathbf{u}_{k-1}^{\mathsf{T}} \mathbf{u}_{k-1});$$
$$\mathbf{u}_k := \frac{\mathbf{Y}_k \mathbf{c}_k}{\|\mathbf{c}_k\|}; \quad \mathbf{c}_k := \mathbf{Y}_k^{\mathsf{T}} \mathbf{t}_k / (\mathbf{t}_k^{\mathsf{T}} \mathbf{t}_k).$$

обладают максимальной ковариацией.

Теорема (Исаченко, 2017)

Оптимальные параметры для линейной модели проекции в скрытое пространство имеют вид

$$\boldsymbol{\Theta} = \mathbf{W}(\mathbf{P}^{\mathsf{T}}\mathbf{W})^{-1}\mathbf{B}\mathbf{Q}^{\mathsf{T}}, \quad \mathbf{T} = \mathbf{X}\mathbf{W}^*, \quad \text{где } \mathbf{W}^* = \mathbf{W}(\mathbf{P}^{\mathsf{T}}\mathbf{W})^{-1}.$$
$$\mathbf{Y} = \mathbf{U}\mathbf{Q}^{\mathsf{T}} + \mathbf{E}_{\mathbf{y}} \approx \mathbf{T}\mathbf{B}\mathbf{Q}^{\mathsf{T}} + \mathbf{E}_{\mathbf{y}} = \mathbf{X}\mathbf{W}^*\mathbf{B}\mathbf{Q}^{\mathsf{T}} + \mathbf{E} = \mathbf{X}\boldsymbol{\Theta} + \mathbf{E}_{\mathbf{y}}.$$

Модель декодирования сигналов

Пусть $f_1(\mathbf{x}_1, \boldsymbol{\Theta}_1)$, $f_2(\mathbf{x}_2, \boldsymbol{\Theta}_2)$ – линейные регрессионные модели декодирования сигналов.

Утверждение (Исаченко, 2021)

Пусть модель декодирования является аддитивной суперпозицией линейных моделей:

$$\mathbf{y} = f_1(\mathbf{x}_1, \boldsymbol{\Theta}_1) + f_2(\mathbf{x}_2, \boldsymbol{\Theta}_2) + \varepsilon = \boldsymbol{\Theta}_1 \mathbf{x}_1 + \boldsymbol{\Theta}_2 \mathbf{x}_2 + \varepsilon.$$

Тогда оптимальные параметры имеют вид

$$\boldsymbol{\Theta}_1 = (\mathbf{X}_1^{\mathsf{T}} \mathbf{M}_{\mathbf{X}_2} \mathbf{X}_1)^{-1} \mathbf{X}_1^{\mathsf{T}} \mathbf{M}_{\mathbf{X}_2} \mathbf{Y}$$
$$\boldsymbol{\Theta}_2 = (\mathbf{X}_2^{\mathsf{T}} \mathbf{M}_{\mathbf{X}_1} \mathbf{X}_2)^{-1} \mathbf{X}_2^{\mathsf{T}} \mathbf{M}_{\mathbf{X}_1} \mathbf{Y},$$

где $\mathbf{M}_{\mathbf{X}_1} = \mathbf{I} - \mathbf{X}_1 (\mathbf{X}_1^{\mathsf{T}} \mathbf{X}_1)^{-1} \mathbf{X}_1^{\mathsf{T}}$, $\mathbf{M}_{\mathbf{X}_2} = \mathbf{I} - \mathbf{X}_2 (\mathbf{X}_2^{\mathsf{T}} \mathbf{X}_2)^{-1} \mathbf{X}_2^{\mathsf{T}}$.

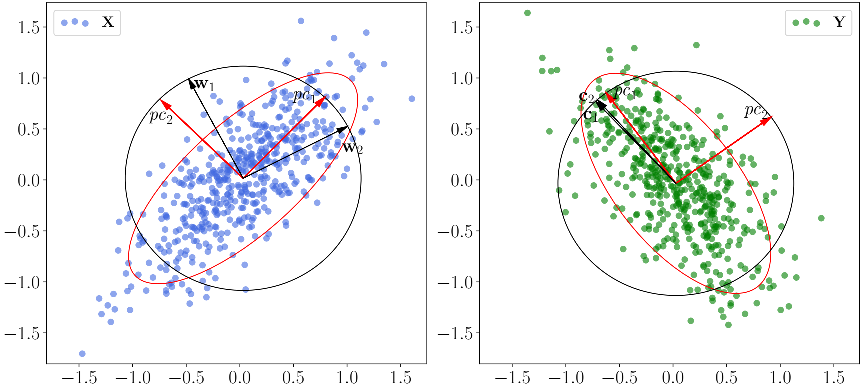
Теорема (Исаченко, 2021)

Если $span(\mathbf{X}_1) \neq span(\mathbf{X}_2)$, *то ошибка аддитивной суперпозиции линейных моделей декодирования не превышает ошибки отдельной модели:*

$$\mathcal{L}(\boldsymbol{\Theta}_1^*, \boldsymbol{\Theta}_2^*, \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}) \leqslant \mathcal{L}(f_i, \mathbf{X}_i, \mathbf{Y}), \quad i = 1, 2.$$

Пример проекции в скрытое пространство в двумерном случае

- $\mathbf{x}_i \sim \mathcal{N}(0, \mathbf{\Sigma})$;
- $\mathbf{y}_i$ линейно зависят от pc_2 и не зависят от pc_1.



Учёт взаимной корреляции между проекциями матриц $\mathbf{X}$ и $\mathbf{Y}$ отклоняет вектора $\mathbf{w}_k$ и $\mathbf{c}_k$ от направления главных компонент.

Скрытое пространство в задаче декодирования

Особенностью задачи является избыточность описания переменных $\mathbf{x}$ и $\mathbf{y}$. Объекты $\mathbf{x}$ и $\mathbf{y}$ живут на некоторых многообразиях низкой размерности.

$$\begin{tikzcd}
\mathbb{X} \subset \mathbb{R}^n \arrow[r, "f"] \arrow[d, shift right, "\varphi_e"'] & \mathbb{Y} \subset \mathbb{R}^r \arrow[d, shift left, "\psi_e"] \\
\mathbb{T} \subset \mathbb{R}^l \arrow[r, "h"] \arrow[u, shift right, "\varphi_d"'] & \mathbb{U} \subset \mathbb{R}^s \arrow[u, shift left, "\psi_d"]
\end{tikzcd}$$

Пространства $\mathbb{T} \subset \mathbb{R}^l$ и $\mathbb{U} \subset \mathbb{R}^s$ *скрытые пространства* для $\mathbb{X} \in \mathbb{R}^n$ $(l \leqslant n)$ и $\mathbb{Y} \in \mathbb{R}^r (s \leqslant r)$, если существуют *функции кодировании* $\varphi_e : \mathbb{X} \to \mathbb{T}$, $\psi_e : \mathbb{Y} \to \mathbb{U}$ и *функции декодирования* $\varphi_d : \mathbb{T} \to \mathbb{X}$, $\psi_d : \mathbb{U} \to \mathbb{Y}$:

$$\begin{aligned}
&\text{для любого } \mathbf{x} \in \mathbb{X} \quad \text{существует } \mathbf{t} \in \mathbb{T} : \varphi_d(\varphi_e(\mathbf{x})) = \varphi_d(\mathbf{t}) = \mathbf{x}; \\
&\text{для любого } \mathbf{y} \in \mathbb{Y} \quad \text{существует } \mathbf{u} \in \mathbb{U} : \psi_d(\psi_e(\mathbf{y})) = \psi_d(\mathbf{u}) = \mathbf{y}.
\end{aligned}$$

Скрытые пространства $\mathbb{T}$ и $\mathbb{U}$ являются *согласованными*, если существует *функция согласования* $h : \mathbb{T} \to \mathbb{U}$:

$$\mathbf{y} = f(\mathbf{x}) = \psi_d(h(\varphi_e(x))).$$

Нелинейные методы согласования скрытого пространства

Функции кодирования и декодирования являются глубокими нейросетями вида:

$$
\begin{aligned}
\mathbf{T} &= \varphi_e(\mathbf{X}) = \mathbf{W}_x^L \sigma(\dots \sigma(\mathbf{W}_x^2 \sigma(\mathbf{X}\mathbf{W}_x^1))\dots)\\
\mathbf{U} &= \psi_e(\mathbf{Y}) = \mathbf{W}_y^L \sigma(\dots \sigma(\mathbf{W}_y^2 \sigma(\mathbf{Y}\mathbf{W}_y^1))\dots)\\
\mathbf{X} &= \varphi_d(\mathbf{X}) = \mathbf{W}_t^L \sigma(\dots \sigma(\mathbf{W}_x^2 \sigma(\mathbf{T}\mathbf{W}_t^1))\dots)\\
\mathbf{Y} &= \psi_d(\mathbf{Y}) = \mathbf{W}_u^L \sigma(\dots \sigma(\mathbf{W}_y^2 \sigma(\mathbf{U}\mathbf{W}_u^1))\dots)
\end{aligned}
$$

$$
\begin{tikzcd}
\mathbb{X} \subset \mathbb{R}^n \arrow[r, "f"] \arrow[d, shift right, "\varphi_e"'] & \mathbb{Y} \subset \mathbb{R}^r \arrow[d, shift left, "\psi_e"] \\
\mathbb{T} \subset \mathbb{R}^l \arrow[r, "h"] \arrow[u, shift right, "\varphi_d"'] & \mathbb{U} \subset \mathbb{R}^s \arrow[u, shift left, "\psi_d"]
\end{tikzcd}
$$

Согласование проекций

Для нахождения оптимальной модели декодирования предложен метод согласования нелинейных проекций

$$
g(\mathbf{T}, \mathbf{U}) \to \max_{\mathbf{W}}, \quad \mathbf{W} = \{\{\mathbf{W}_x^i\}_{i=1}^L, \{\mathbf{W}_y^i\}_{i=1}^L, \{\mathbf{W}_t^i\}_{i=1}^L, \{\mathbf{W}_u^i\}_{i=1}^L\}.
$$

Выбор признаков в задаче декодирования

Требуется

Найти бинарный вектор $\mathbf{a} = \{0, 1\}^n$, компоненты – индикаторы выбранных признаков.

Функция ошибки отбора признаков

$$\mathbf{a} = \underset{\mathbf{a}' \in \{0,1\}^n}{\arg\min}\ S(\mathbf{a}' | \mathbf{X}, \mathbf{Y}).$$

Релаксация

Замена дикретной области определения $\{0, 1\}^n$ на непрерывную релаксацию $[0, 1]^n$:

$$\mathbf{z} = \underset{\mathbf{z}' \in [0,1]^n}{\arg\min}\ S(\mathbf{z}' | \mathbf{X}, \mathbf{Y}), \quad a_j = [z_j > \tau].$$

Получив $\mathbf{a}$, решаем задачу регрессии:

$$\mathcal{L}(\mathbf{\Theta_a} | \mathbf{X_a}, \mathbf{Y}) = \left\| \mathbf{Y} - \mathbf{X_a} \mathbf{\Theta}_\mathbf{a}^\mathsf{T} \right\|_2^2 \to \min_{\mathbf{\Theta_a}},$$

где индекс $\mathbf{a}$ обозначает подматрицу с номерами столбцов, для которых $a_j = 1$.

Выбор признаков с помощью квадратичного программирования

$\mathbf{X} = [\boldsymbol{\chi}_1, \dots, \boldsymbol{\chi}_n] \in \mathbb{R}^{m \times n}$ – матрица объектов; $\mathbf{Y} = [\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_r] \in \mathbb{R}^{m \times r}$ – матрица ответов.

$$\|\boldsymbol{\nu} - \mathbf{X}\boldsymbol{\theta}\|_2^2 \to \min_{\boldsymbol{\theta} \in \mathbb{R}^n}.$$

Задача квадратичного программирования

$$S(\mathbf{z}, \mathbf{Q}, \mathbf{b}) = (1 - \alpha) \cdot \underbrace{\mathbf{z}^{\mathsf{T}} \mathbf{Q} \mathbf{z}}_{\text{Sim}(\mathbf{X})} - \alpha \cdot \underbrace{\mathbf{b}^{\mathsf{T}} \mathbf{z}}_{\text{Rel}(\mathbf{X}, \boldsymbol{\nu})} \to \min_{\substack{\mathbf{z} \geqslant \mathbf{0}_n \\ \mathbf{1}_n^{\mathsf{T}} \mathbf{z} = 1}}.$$

$\mathbf{z} \in [0, 1]^n$ – значимость признаков;
$\mathbf{Q} = \left[|\text{corr}(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j)|\right]_{i,j=1}^n \in \mathbb{R}^{n \times n}$ – матрица парных взаимодействий признаков;
$\mathbf{b} = \left[|\text{corr}(\boldsymbol{\chi}_i, \boldsymbol{\nu})|\right]_{i=1}^n \in \mathbb{R}^n$ – вектор релевантностей признаков к целевой переменной.

Утверждение (Исаченко, 2018)

Задача $S(\mathbf{z}, \hat{\mathbf{Q}}, \mathbf{b})$ имеет единственный глобальный минимум, где матрица $\hat{\mathbf{Q}}$ получена полуопределенной релаксацией исходной матрицы $\mathbf{Q}$:

$$\hat{\mathbf{Q}} = \mathbf{Q} - \lambda_{\min}(\mathbf{Q}) \mathbf{I}.$$

Многомерный выбор признаков в задаче декодирования

Агрегирование релевантностей по целевым векторам (RelAgg)

$$\mathbf{b} = \left[|\text{corr}(\boldsymbol{\chi}_i, \boldsymbol{\nu})|\right]_{i=1}^{n} \to \mathbf{b} = \left[\sum_{k=1}^{r} |\text{corr}(\boldsymbol{\chi}_i, \boldsymbol{\nu}_k)|\right]_{i=1}^{n}.$$

Недостаток: нет учёта зависимостей в матрице $\mathbf{Y}$.

Симметричный учёт значимостей (SymImp)

Штрафуем коррелированные целевые вектора с помощью $\text{Sim}(\mathbf{Y})$

$$\alpha_1 \cdot \underbrace{\mathbf{z}_x^{\mathsf{T}} \mathbf{Q}_x \mathbf{z}_x}_{\text{Sim}(\mathbf{X})} - \alpha_2 \cdot \underbrace{\mathbf{z}_x^{\mathsf{T}} \mathbf{B} \mathbf{z}_y}_{\text{Rel}(\mathbf{X}, \mathbf{Y})} + \alpha_3 \cdot \underbrace{\mathbf{z}_y^{\mathsf{T}} \mathbf{Q}_y \mathbf{z}_y}_{\text{Sim}(\mathbf{Y})} \to \min_{\substack{\mathbf{z}_x \geqslant \mathbf{0}_n,\, \mathbf{1}_n^{\mathsf{T}} \mathbf{z}_x = 1 \\ \mathbf{z}_y \geqslant \mathbf{0}_r,\, \mathbf{1}_r^{\mathsf{T}} \mathbf{z}_y = 1}}.$$

$$\mathbf{Q}_x = \left[|\text{corr}(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j)|\right]_{i,j=1}^{n}, \ \mathbf{Q}_y = \left[|\text{corr}(\boldsymbol{\nu}_i, \boldsymbol{\nu}_j)|\right]_{i,j=1}^{r}, \ \mathbf{B} = \left[|\text{corr}(\boldsymbol{\chi}_i, \boldsymbol{\nu}_j)|\right]_{\substack{i=1,\dots,n \\ j=1,\dots,r}}.$$

$$\alpha_1 + \alpha_2 + \alpha_3 = 1, \quad \alpha_i \geqslant 0.$$

Многомерный выбор признаков в задаче декодирования

SymImp штрафует коррелированные целевые вектора, которые в меньшей мере объясняются признаками.

$$\alpha_1 \cdot \underbrace{\mathbf{z}_x^{\mathsf{T}} \mathbf{Q}_x \mathbf{z}_x}_{\mathrm{Sim}(\mathbf{X})} - \alpha_2 \cdot \underbrace{\mathbf{z}_x^{\mathsf{T}} \mathbf{B} \mathbf{z}_y}_{\mathrm{Rel}(\mathbf{X},\mathbf{Y})} \to \min_{\substack{\mathbf{z}_x \geqslant \mathbf{0}_n \\ \mathbf{1}_n^{\mathsf{T}} \mathbf{z}_x = 1}}; \quad \alpha_3 \cdot \underbrace{\mathbf{z}_y^{\mathsf{T}} \mathbf{Q}_y \mathbf{z}_y}_{\mathrm{Sim}(\mathbf{Y})} + \alpha_2 \cdot \underbrace{\mathbf{z}_x^{\mathsf{T}} \mathbf{B} \mathbf{z}_y}_{\mathrm{Rel}(\mathbf{X},\mathbf{Y})} \to \min_{\substack{\mathbf{z}_y \geqslant \mathbf{0}_r \\ \mathbf{1}_r^{\mathsf{T}} \mathbf{z}_y = 1}}.$$

Минимаксный подход (MinMax / MaxMin)

$$\min_{\substack{\mathbf{z}_x \geqslant \mathbf{0}_n \\ \mathbf{1}_n^{\mathsf{T}} \mathbf{z}_x = 1}} \max_{\substack{\mathbf{z}_y \geqslant \mathbf{0}_r \\ \mathbf{1}_r^{\mathsf{T}} \mathbf{z}_y = 1}} \left(\text{or} \max_{\substack{\mathbf{z}_y \geqslant \mathbf{0}_r \\ \mathbf{1}_r^{\mathsf{T}} \mathbf{z}_y = 1}} \min_{\substack{\mathbf{z}_x \geqslant \mathbf{0}_n \\ \mathbf{1}_n^{\mathsf{T}} \mathbf{z}_x = 1}} \right) \left[\alpha_1 \cdot \underbrace{\mathbf{z}_x^{\mathsf{T}} \mathbf{Q}_x \mathbf{z}_x}_{\mathrm{Sim}(\mathbf{X})} - \alpha_2 \cdot \underbrace{\mathbf{z}_x^{\mathsf{T}} \mathbf{B} \mathbf{z}_y}_{\mathrm{Rel}(\mathbf{X},\mathbf{Y})} - \alpha_3 \cdot \underbrace{\mathbf{z}_y^{\mathsf{T}} \mathbf{Q}_y \mathbf{z}_y}_{\mathrm{Sim}(\mathbf{Y})} \right].$$

Теорема (Исаченко, 2018)

Для положительно определенных матриц $\mathbf{Q}_x$ и $\mathbf{Q}_y$ minmax и maxmin задачи достигают одинакового значения функционала.

Теорема (Исаченко, 2018)

Минимаксная задача эквивалентна задаче квадратичного программирования с $n + r + 1$ переменными.

Для получения выпуклой задачи применяется полуопределенная рекласация сдвига спектра.

Многомерный выбор признаков в задаче декодирования

Максимизация релевантностей (MaxRel)

$$\min_{\substack{\mathbf{z}_x \geqslant \mathbf{0}_n \\ \mathbf{1}_n^{\mathsf{T}} \mathbf{z}_x = 1}} \max_{\substack{\mathbf{z}_y \geqslant \mathbf{0}_r \\ \mathbf{1}_r^{\mathsf{T}} \mathbf{z}_y = 1}} \left[(1-\alpha) \cdot \mathbf{z}_x^{\mathsf{T}} \mathbf{Q}_x \mathbf{z}_x - \alpha \cdot \mathbf{z}_x^{\mathsf{T}} \mathbf{B} \mathbf{z}_y \right].$$

Теорема (Исаченко, 2018)

Для положительно определенной матрицы $\mathbf{Q}_x$ minmax и maxmin задачи достигают одинакового значения функционала.

Асимметричный учёт значимостей (AsymImp)

$$\alpha_1 \cdot \underbrace{\mathbf{z}_x^{\mathsf{T}} \mathbf{Q}_x \mathbf{z}_x}_{\mathrm{Sim}(\mathbf{X})} - \alpha_2 \cdot \underbrace{\left(\mathbf{z}_x^{\mathsf{T}} \mathbf{B} \mathbf{z}_y - \mathbf{b}^{\mathsf{T}} \mathbf{z}_y \right)}_{\mathrm{Rel}(\mathbf{X}, \mathbf{Y})} + \alpha_3 \cdot \underbrace{\mathbf{z}_y^{\mathsf{T}} \mathbf{Q}_y \mathbf{z}_y}_{\mathrm{Sim}(\mathbf{Y})} \to \min_{\substack{\mathbf{z}_x \geqslant \mathbf{0}_n,\, \mathbf{1}_n^{\mathsf{T}} \mathbf{z}_x = 1 \\ \mathbf{z}_y \geqslant \mathbf{0}_r,\, \mathbf{1}_r^{\mathsf{T}} \mathbf{z}_y = 1}}.$$

При $b_j = \max\limits_{i=1,\dots n} [\mathbf{B}]_{i,j}$ коэффициенты при $\mathbf{z}_y$ в $\mathrm{Rel}(\mathbf{X}, \mathbf{Y})$ неотрицательны.

Обобщение предложенных методов выбора признаков

Алгоритм	Критерий	Функция ошибки $S(\mathbf{z}, \mathbf{X}, \mathbf{Y})$
RelAgg	$\min \bigl[\mathrm{Sim}(\mathbf{X}) - \mathrm{Rel}(\mathbf{X}, \mathbf{Y})\bigr]$	$\min\limits_{\mathbf{z}_x} \bigl[(1-\alpha) \cdot \mathbf{z}_x^{\mathsf{T}} \mathbf{Q}_x \mathbf{z}_x - \alpha \cdot \mathbf{z}_x^{\mathsf{T}} \mathbf{B} \mathbf{1}_r\bigr]$
SymImp	$\min \bigl[\mathrm{Sim}(\mathbf{X}) - \mathrm{Rel}(\mathbf{X}, \mathbf{Y}) + \mathrm{Sim}(\mathbf{Y})\bigr]$	$\min\limits_{\mathbf{z}_x, \mathbf{z}_y} \bigl[\alpha_1 \cdot \mathbf{z}_x^{\mathsf{T}} \mathbf{Q}_x \mathbf{z}_x - \alpha_2 \cdot \mathbf{z}_x^{\mathsf{T}} \mathbf{B} \mathbf{z}_y + \alpha_3 \cdot \mathbf{z}_y^{\mathsf{T}} \mathbf{Q}_y \mathbf{z}_y\bigr]$
MinMax	$\min \bigl[\mathrm{Sim}(\mathbf{X}) - \mathrm{Rel}(\mathbf{X}, \mathbf{Y})\bigr]$ $\max \bigl[\mathrm{Rel}(\mathbf{X}, \mathbf{Y}) + \mathrm{Sim}(\mathbf{Y})\bigr]$	$\min\limits_{\mathbf{z}_x} \max\limits_{\mathbf{z}_y} \bigl[\alpha_1 \cdot \mathbf{z}_x^{\mathsf{T}} \mathbf{Q}_x \mathbf{z}_x - \alpha_2 \cdot \mathbf{z}_x^{\mathsf{T}} \mathbf{B} \mathbf{z}_y - \alpha_3 \cdot \mathbf{z}_y^{\mathsf{T}} \mathbf{Q}_y \mathbf{z}_y\bigr]$
MaxRel	$\min \bigl[\mathrm{Sim}(\mathbf{X}) - \mathrm{Rel}(\mathbf{X}, \mathbf{Y})\bigr]$ $\max \bigl[\mathrm{Rel}(\mathbf{X}, \mathbf{Y})\bigr]$	$\min\limits_{\mathbf{z}_x} \max\limits_{\mathbf{z}_y} \bigl[(1-\alpha) \cdot \mathbf{z}_x^{\mathsf{T}} \mathbf{Q}_x \mathbf{z}_x - \alpha \cdot \mathbf{z}_x^{\mathsf{T}} \mathbf{B} \mathbf{z}_y\bigr]$
AsymImp	$\min \bigl[\mathrm{Sim}(\mathbf{X}) - \mathrm{Rel}(\mathbf{X}, \mathbf{Y})\bigr]$ $\max \bigl[\mathrm{Rel}(\mathbf{X}, \mathbf{Y}) + \mathrm{Sim}(\mathbf{Y})\bigr]$	$\min\limits_{\mathbf{z}_x, \mathbf{z}_y} \bigl[\alpha_1 \mathbf{z}_x^{\mathsf{T}} \mathbf{Q}_x \mathbf{z}_x - \alpha_2 \left(\mathbf{z}_x^{\mathsf{T}} \mathbf{B} \mathbf{z}_y - \mathbf{b}^{\mathsf{T}} \mathbf{z}_y\right) + \alpha_3 \mathbf{z}_y^{\mathsf{T}} \mathbf{Q}_y \mathbf{z}_y\bigr]$

Теорема (Исаченко, 2018)

В одномерном случае $r = 1$ предлагаемые методы выбора признаков SymImp, MinMax, MaxMin, MaxRel, AsymImp совпадают с исходной задачей минимизации функции ошибок $S(\mathbf{z}, \mathbf{Q}, \mathbf{b})$.

Внешние критерии качества решения задачи декодирования

Нормированное RMSE

Качество предсказания:

$$\text{sRMSE}(\mathbf{Y}, \widehat{\mathbf{Y}}_\mathbf{a}) = \sqrt{\frac{\text{MSE}(\mathbf{Y}, \widehat{\mathbf{Y}}_\mathbf{a})}{\text{MSE}(\mathbf{Y}, \overline{\mathbf{Y}})}} = \frac{\|\mathbf{Y} - \widehat{\mathbf{Y}}_\mathbf{a}\|_2}{\|\mathbf{Y} - \overline{\mathbf{Y}}\|_2}, \quad \text{где} \quad \widehat{\mathbf{Y}}_\mathbf{a} = \mathbf{X}_\mathbf{a}\mathbf{\Theta}_\mathbf{a}^\mathsf{T}.$$

$\overline{\mathbf{Y}}$ — константный прогноз.

Мультикорреляция

Среднее значение коэффициента множественной корреляции:

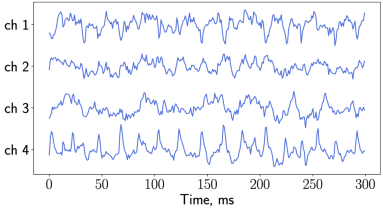
$$R^2 = \frac{1}{r}\text{tr}\left(\mathbf{C}^\mathsf{T}\mathbf{R}^{-1}\mathbf{C}\right); \quad \mathbf{C} = [\text{corr}(\boldsymbol{\chi}_i, \boldsymbol{\nu}_j)]_{\substack{i=1,\dots,n \\ j=1,\dots,r}}, \ \mathbf{R} = [\text{corr}(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j)]_{i,j=1}^n.$$

BIC

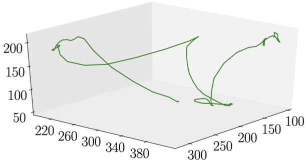
Компромисс между качеством предсказания и числом выбранных признаков $\|\mathbf{a}\|_0$:

$$\text{BIC} = m\ln\left(\text{MSE}(\mathbf{Y}, \widehat{\mathbf{Y}}_\mathbf{a})\right) + \|\mathbf{a}\|_0 \cdot \log m.$$

Задача восстановления траектории конечности по сигналам электрокортикограммы



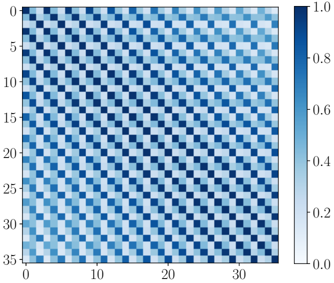
$\xrightarrow{f}$



- $\mathbf{X} \in \mathbb{R}^{m\times(32\cdot 27)}$ – сигналы ECoG.
- $\mathbf{Y} \in \mathbb{R}^{m\times 3k}$ – траектория движения руки.

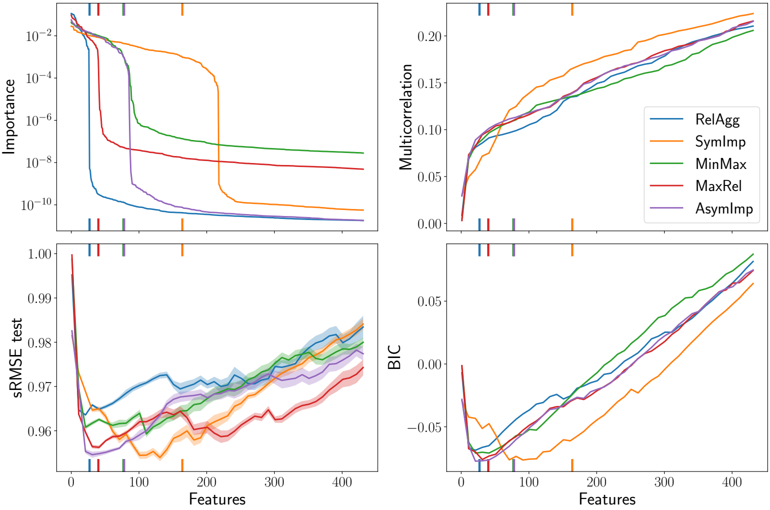
$$\mathbf{Y} = \begin{pmatrix} x_1\ y_1\ z_1 & \dots & x_k & y_k & z_k \\ x_2\ y_2\ z_2 & \dots & x_{k+1} & y_{k+1} & z_{k+1} \\ \dots & \dots & & \dots & \\ x_m\ y_m\ z_m & \dots & x_{m+k} & y_{m+k} & z_{m+k} \end{pmatrix}$$

http://neurotycho.org



Матрица корреляций **Y**

Анализ предложенных методов выбора признаков



Предложенные методы выбирают модель с меньшей ошибкой по отношению к базовому алгоритму.

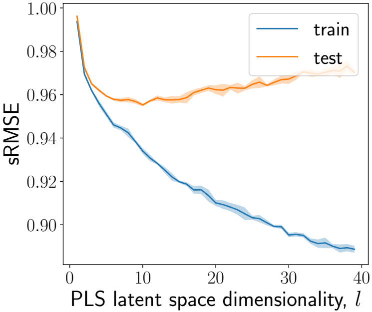
Стабильность методов выбора признаков

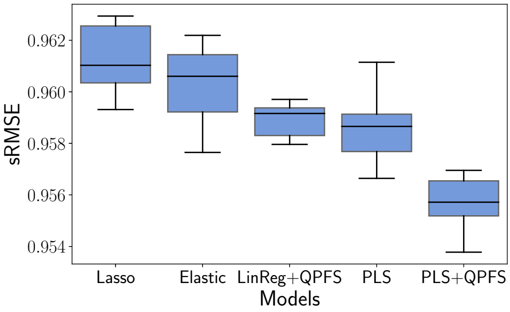
Постановка эксперимента

- создать бутстреп-выборки
$$(\mathbf{X}, \mathbf{Y}) \to \big\{(\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_s, \mathbf{Y}_s)\big\};$$
- решить задачу выбора признаков
$$\big\{(\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_s, \mathbf{Y}_s)\big\} \to \{\mathbf{z}_1, \dots, \mathbf{z}_s\};$$
- вычислить статистики
$$\{\mathbf{z}_1, \dots, \mathbf{z}_s\} \to \{\text{sRMSE}, \|\mathbf{a}\|_0, \text{Спирмен } \rho, \ell_2 \text{ расстояние}\}.$$

	sRMSE	$\|\mathbf{a}\|_0$	Спирмен ρ	ℓ_2 расстояние
RelAgg	0.965 ± 0.002	26.8 ± 3.8	0.915 ± 0.016	0.145 ± 0.018
SymImp	0.961 ± 0.001	224.4 ± 9.0	0.910 ± 0.017	0.025 ± 0.002
MinMax	0.961 ± 0.002	101.0 ± 2.1	0.932 ± 0.009	0.059 ± 0.004
MaxRel	0.958 ± 0.003	41.2 ± 5.2	0.862 ± 0.027	0.178 ± 0.010
AsymImp	0.955 ± 0.001	85.8 ± 10.2	0.926 ± 0.011	0.078 ± 0.007

Сравнение метода проекции в скрытое пространство с методами выбора признаков





- Предлагаемые методы выбора признаков достигают меньшей ошибки по сравнению с базовыми алгоритмами Lasso и Elastic.
- PLS показывает сравнимое качество с QPFS.
- Комбинация двух алгоритмов показывает наилучший результат.

Результаты, выносимые на защиту

1. Исследована задача декодирования сигналов в пространствах высокой размерности. Исследованы методы снижения размерности с анализом структуры пространства.

2. Доказаны теоремы об оптимальности предлагаемых методов декодирования сигналов.

3. Предложены методы для выбора признаков, учитывающие зависимости как в пространстве объектов, так и в пространстве ответов. Предложенные алгоритмы доставляют устойчивые и адекватные решения в коррелированных пространствах высокой размерности.

4. Предложены нелинейные методы согласования скрытых пространств для данных со сложно организованной целевой переменной.

5. Создан макет системы решения задачи декодирования сигналов в пространствах высокой размерности.

Заключение

Публикации ВАК

1. Isachenko R., Strijov V. Quadratic Programming Feature Selection for Multicorrelated Signal Decoding with Partial Least Squares *Expert Systems with Applications*, 2021, на рецензировании.
2. Исаченко Р.В., Яушев Ф.Р., Стрижов В.В. Модели согласования скрытого пространства в задаче прогнозирования // Системы и средства информатики, 31(1), 2021.
3. Isachenko R., Vladimirova M., Strijov V. Dimensionality Reduction for Time Series Decoding and Forecasting Problems. *DEStech Transactions on Computer Science and Engineering*, optim, 2018.
4. Isachenko R., Strijov V. Quadratic programming optimization for Newton method. *Lobachevskii Journal of Mathematics*, 39(9), 2018.
5. Isachenko R. et al. Feature Generation for Physical Activity Classification. *Artificial Intellegence and Decision Making*, 3, 2018.
6. Исаченко Р.В., Стрижов В. В. Метрическое обучение в задачах мультиклассовой классификации временных рядов *Информатика и её применения*, 10(2), 2016.

Выступления с докладом

1. Intelligent Data Processing Conference, 2020, Снижение размерности в задаче декодирования временных рядов.
2. Intelligent Data Processing Conference, 2018, Dimensionality reduction for multicorrelated signal decoding with projections to latent space.
3. Математические методы распознавания образов, 2017. Локальные модели для классификации объектов сложной структуры.
4. Intelligent Data Processing Conference, 2016. Multimodel forecasting multiscale time series in internet of things.
5. Ломоносов, 2016. Метрическое обучение в задачах мультиклассовой классификации временных рядов.